FEBS Letters 584 (2010) 4041–4047

FEBS Letters

journal homepage: www.FEBSLetters.org

Alternative splicing of genes during neuronal differentiation of NT2 pluripotential human embryonal carcinoma cells

Ai Wakamatsu [a], Jun-ichi Imai [b], Shinya Watanabe [b], Takao Isogai [a,c,*]

[a] Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan
[b] Department of Clinical Genomics, Translational Research Center (Tokyo Branch), Fukushima Medical University, Shibuya-ku, Tokyo, Japan
[c] Reverse Proteomics Research Institute, Chiyoda-ku, Tokyo, Japan

ARTICLE INFO

Article history:
Received 13 July 2010
Revised 11 August 2010
Accepted 13 August 2010
Available online 20 August 2010

Edited by Takashi Gojobori

Keywords:
Alternative splicing
Neuronal differentiation
mRNA diversity
Transcription factor
Expression profile
Multiple protein-coding transcript

ABSTRACT

We analyzed the mRNA diversity of genes after inducing neuronal differentiation in human NT2 teratocarcinoma cells using all-trans retinoic acid (RA). DNA microarray analyses of cells treated with RA identified 358 RA-responsive genes. mRNA diversity analysis revealed that 274 genes produced multiple protein-coding transcripts by alternative splicing. Among these 274 genes, we chose 26 genes that showed AS in their C-terminus and 12 transcription factor genes for further analysis. By using transcript-specific primers, we performed quantitative real-time PCR analysis to examine the expression profiles of all the protein-coding transcripts. Consequently, we identified genes which showed different RA-induced changes in the expression of their protein-coding transcripts.

1. Introduction

One of the most important prerequisites for understanding the functions of genes is to analyze mRNA diversity. Several genome research and high-throughput sequencing studies suggest that at least 70% of all human genes produce multiple transcripts by alternative splicing (AS) [1,2]. Accordingly, AS of a single gene could produce multiple protein-coding transcripts, each exhibiting different properties, intracellular localization and regulation of enzymatic activities [3]. Several important events during neuronal development, including aspects of cell-fate determination, axon guidance and synaptogenesis, are controlled by AS [4], suggesting that AS might be prevalent in the mammalian nervous system. For example, a switch from PTB to nPTB expression during neurodifferentiation has been implicated in regulating neuronal AS [5]. Various studies have also identified brain-specific splicing factors, such as NOVA1 that regulates neuron-specific AS and is essential for neuronal viability [6,7]. Obviously, such information on mRNA diversity would be important for understanding the mechanism of neurodifferentiation, and could ultimately provide therapeutic solutions for neural degenerative diseases, such as Parkinson's and Alzheimer's diseases. Although the importance of analyzing mRNA diversity in the nervous-system development and function is becoming apparent, the mechanism controlling the neuronal AS has remained unclear. Previously, we sequenced about 55 thousand human full-length cDNAs and about 1.45 million 5′-end sequences (5′-EST) from human cDNA libraries constructed by oligo-capping method [8–11] for gene function analysis. Our previous results suggested a close relationship between the predicted function of a gene and its tissue-specific expression [11]. We also constructed the FLJ Human cDNA Database (http://flj.lifesciencedb.jp) [11], which would be very useful in analyzing the diversity of protein-coding transcripts. Additionally, we found a close relationship between the predicted function of a gene and its

Abbreviations: AS, alternative splicing; CTHRC1, collagen triple helix repeat containing 1; DENND5B, DENN/MADD domain containing 5B; EPB41L5, erythrocyte membrane protein band 4.1 like 5; ETV1, ets variant 1; ETV4, ets variant 4; ETV5, ets variant gene 5; FEV, first exon variation; GAPDH, glyceraldehyde-3-phosphate dehydrogenase; GATA6, GATA binding protein 6; GBX2, gastrulation brain homeobox 2; GO, Gene Ontology; HESX1, HESX homeobox 1; HNF1B, HNF1 homeobox B; HOXA2, homeobox A2; HOXA3, homeobox A3; HOXB2, homeobox B2; HOXB4, homeobox B4; HOXC4, homeobox C4; NPHP1, nephronophthisis 1 (juvenile); NRP2, neuropilin 2; PAX6, paired box 6; PEG3, paternally expressed 3; POU5F1, POU class 5 homeobox 1; RA, all-trans retinoic acid; RARB, retinoic acid receptor, beta; RFX2, regulatory factor X, 2 (influences HLA class II expression); TSS, transcription start site; UTR, untranslated region; WDR74, WD repeat domain 74; ZNF483, zinc finger protein 483

* Corresponding author at: Graduate School of Pharmaceutical Sciences, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan. Fax: +81 3 5841 4775.
E-mail address: tisogai@mol.f.u-tokyo.ac.jp (T. Isogai).

0014-5793/$36.00
doi:10.1016/j.febslet.2010.08.024

tissue-specific expression, implying that the gene was transcribed into an appropriate transcript according to the need and circumstance. However, in our previous study [11], we were unable to distinguish whether the observed changes in the gene expression pattern was associated with the genetic or the environmental risk factors, primarily because the tissues used in our previous analyses were not derived from one individual. In the present study, we used NT2 cells, also known as NT2/D1, pluripotent human embryonal carcinoma cell lines, for analyzing mRNA diversity. This cell line is a good model for the embryonic development studies as well as for the tumor cell differentiation studies. In response to all-trans retinoic acid (RA), NT2 cells differentiate towards a neuronal lineage with associated loss of cell growth and tumorigenicity [12]. Because these cells shared same genetic background, we were able to analyze changes in the mRNA diversity of genes following changes in an environmental condition, namely, stimulation with RA. Thus, our efforts to identify RA-responsive genes and to analyze mRNA diversity have led to various findings on neurodifferentiation.

Here, we have used the DNA microarray technique to globally analyze genes in the NT2 cells whose expression levels changed when cells were treated with RA. We have also used the FLJ Human cDNA Database to explore correlations between the mRNA diversity of gene and gene function. Finally, we have identified the protein-coding transcripts generated from selected each one of the selected 38 genes, and examined their expression profiles by real-time PCR analysis.

2. Materials and methods

2.1. Cell culture and RNA isolation

NT2 cell line was purchased from STRATAGENE (CA, USA) [12] and was cultured according to the recommendations of the supplier. The stock solution (10 mM) of RA (SIGMA, MO, USA) was prepared in dimethyl sulfoxide. On day 0, the culture medium was replaced with fresh medium containing 10 μM RA. After 24, 48, and 168 h incubation with RA, we collected the cells for RNA isolation and designated them as 1-day, 2-day and 7-day samples, respectively. The 0-day sample (cells immediately before the addition of RA) was used as the control. Total RNAs were extracted from these cells using the ISOGEN Reagent (Nippongene, Tokyo, Japan). Subsequently, polyA(+) RNAs were isolated from each total RNA sample using a MicroPoly(A) Purist Kit (Life Technologies, CA, USA).

2.2. Microarray analysis

Synthetic polynucleotides (80-mers), representing 31 917 human transcripts, (MicroDiagnostic, Tokyo, Japan) were arrayed by using a custom arrayer. Microarray analyses were performed as previously described [13], and detailed method is described in the Supplementary method 1. All the data in accordance with the MIAME guideline were deposited at DDBJ via CIBEX database (http://cibex.nig.ac.jp/index.jsp) in Accession Numbers CBX132. To identify genes demonstrating significant changes in expression, *t*-test was performed between the 0-day sample (negative control) and each RA time point sample ($P < 0.01$). Among the extracted genes, we further selected those genes that exhibited differences greater than 1.0 between the mean averages of log ratios for the two sample groups.

2.3. Analysis of splicing patterns and functional classification of cDNAs by gene ontology (GO)

Sequences of proteins encoded by the transcripts of 358 genes selected in this paper and 24 210 human RefSeq were categorized by GO (http://www.geneontology.org/), detailed method for GO classification is described in the Supplementary method 2. The splicing patterns of the selected genes were analyzed by using the information available in the FLJ Human cDNA Database ver. 3.0 as described previously [11]. For analyzing the N-terminus splicing patterns, we only used our FLJ ESTs constructed by an optimized oligo-capping method, 90% or more of which contained the transcription start site (TSS) [10,11]. In our analysis, we only focused on the protein-coding transcripts, and ignored a lot of non-coding RNAs and mRNAs in which AS occurred only in the untranslated region [14].

2.4. Quantitative real-time PCR analysis

Synthesis of template cDNA, performance of real-time PCR and design of primers were performed as previously described [11], and detailed method is described in the Supplementary method 3. The expression values of individual genes were calculated by comparing their Ct values to that of the control using the RQ software (Life Technologies). And the expression levels of genes were normalized with respect to that of the human glyceraldehyde-3-phosphate dehydrogenase (GAPDH) and were represented in log2 base. Samples were run in triplicates and the data shown are average of three experiments.

3. Results

3.1. Overview of RA-induced alteration in expression of genes in NT2 cells and functional classification of the selected genes by GO

First, to understand the functions of genes during neurodifferentiation, it was necessary to identify the genes whose expression levels altered upon RA stimulation. Based on the results of previous studies [15,16], we reasoned that the expression levels of many genes would alter within 7 days of stimulation with RA. Accordingly, we compared the gene expression profiles of the 0-, 1-, 2- and 7-day RA-stimulated NT2 cells, and consequently obtained 40, 106 and 340 probes (Fig. 1), respectively, whose expression ratios changed over 1.0 in log 2 scale ($P < 0.01$). We eventually obtained a total of 423 probes by merging the expression level difference data for each RA-stimulation time point, and these results are summarized in Supplementary Table 1. In the 7-day RA-stimulated NT2 cells, the expression levels of the known neurodifferentiation markers NEF3 and neuropilin 2 (NRP2) [17,18] increased by ninefold and threefold, respectively, (Supplementary Fig. 1), whereas the expression levels of MAP2 and MAPT (Tau), known as later stage nerve differentiation markers [19], did not change more than 1.5-fold (Supplementary Fig. 2), suggesting that these RA-stimulated cells underwent neuronal differentiation, but were not at the later nerve differentiation stage. Furthermore, the expression level of the novel tumor progression marker adrenomedullin (ADM), which is known to be up-regulated in neuroendocrine carcinomas [20], in the 7-day RA-stimulated cells decreased to one-third of that of the control cells (Supplementary Fig. 1), suggesting that these cells lost their carcinoma phenotype. Thus, by using this experimental system one could identify genes involved in neuronal differentiation and also could determine the underlying cause of a disease. Our results of DNA microarray analyses suggest that a large number of genes might be involved in the neuronal differentiation process. Thus, to understand the mechanism of neuronal differentiation, it is important to analyze functions of these genes and find correlation among them.

Next, it was necessary to identify genome loci of the selected 423 probes. Analysis of these 423 selected probes, as described previously [11], led to the identification of 358 protein-coding

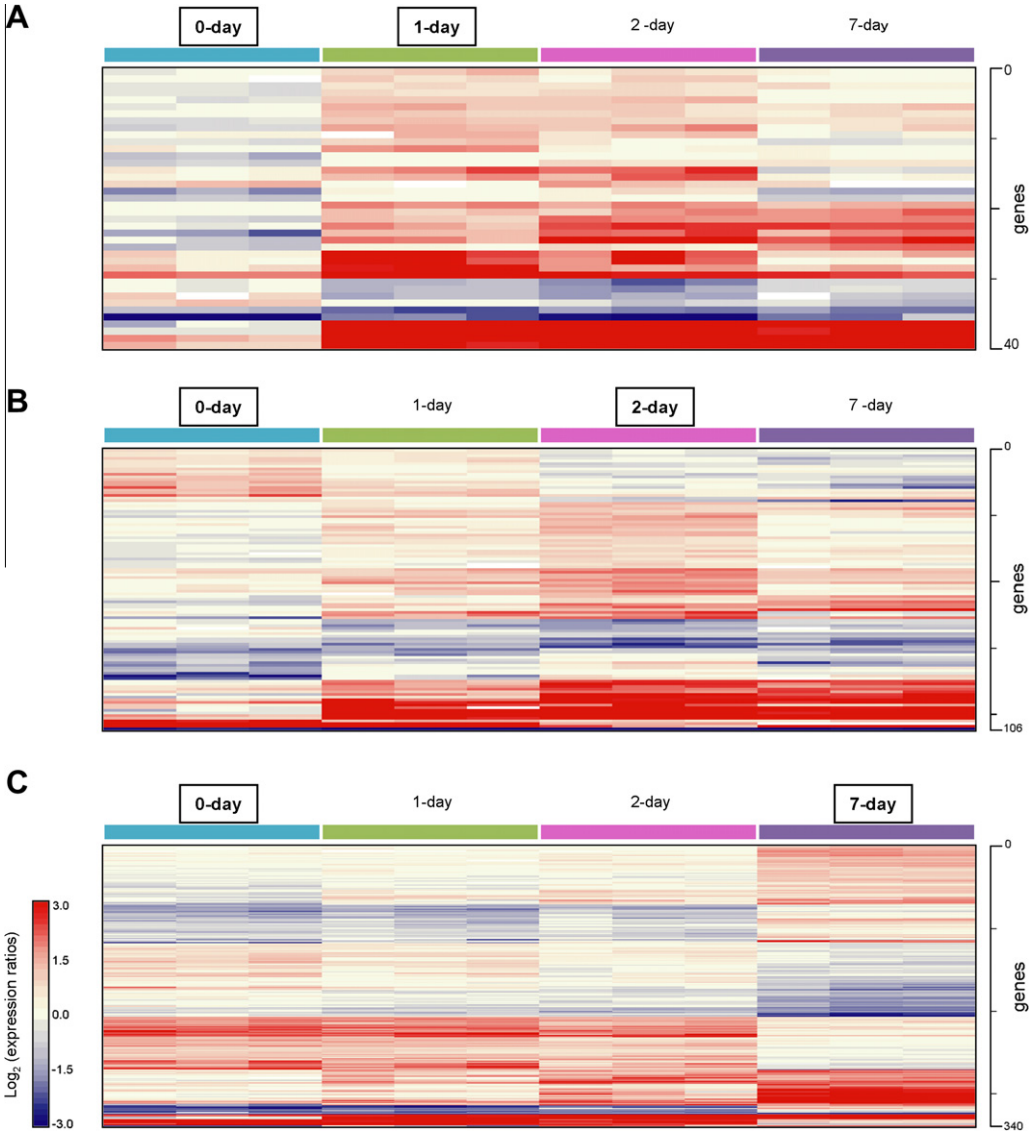

Fig. 1. Comparison of gene expression profiles of control and RA-induced NT2 cells. Gene expression profiles of RA-treated cell data after filtering out the control cell data as described under Materials and methods. (A) 1-day, (B) 2-day and (C) 7-day samples, included 40, 106, 340 probes, respectively. (Lists of genes are shown in Supplementary Table 1). Columns and rows indicate RA time point samples and genes, respectively. Genes and samples are aligned in the order defined by the results of the clustering analysis. The color bar represents the grades of the relative expression levels: increase, red; decrease, blue.

genes (Supplementary Table 2), which were then classified according to GO classification to predict their putative functions. As shown in Table 1, majority of the selected genes were classified under the GO molecular function categories "transcription regulator activity", "ligase activity" and "ion binding". Most genes belonging to the "ligase activity" and "ion binding" categories showed altered expression levels in 7-day sample (Supplementary Table 3), and they are most likely the candiate genes that are involved in the neurodifferentiation. Eighteen genes belonging to the "transcription regulator activity" category were identified as transcription factors. Because expression levels of these transcription factor genes changed in all three RA-stimulated samples (Supplementary Fig. 3), we thought they might regulate expression of other genes related to neurodifferentiation. Therefore, we analyzed them in more detail (see below).

3.2. mRNA diversity of RA-induced genes

We have previously found that many genes show tissue-specific splicing patterns [11]. Therefore, we next analyzed the mRNA diversity in 358 selected genes using our FLJ Human cDNA Database. Consequently, we obtained 274 genes (~77% of selected genes) producing multiple protein-coding transcripts as a result of AS (Supplementary Fig. 4 and Supplementary Table 2).

First, we evaluated the type of splicing using reliable transcripts. To exclude the unreliable transcripts from the immature mRNA and genomic DNA, reliability of mRNA was evaluated by using huge number of the human EST data. As a result, we found 136 genes producing multiple protein-coding transcripts using alternate N-terminus (Alt. N-term) and 59 genes producing multiple protein-coding transcripts using alternate C-terminus (Alt. C-term). Additionally, there were 26 genes producing multiple protein-coding transcripts by using alternate use of both N- and C-terminuses. We have previously found that alternative TSSs were shown to be utilized for tissue-specific expression [10,11], and there were demonstrated relationship between the TSS selectivity and mechanism of transcription for some genes [21,22]. Therefore, we next analyzed the mRNA diversity of 59 genes each one of which produced multiple protein-coding transcripts by using Alt. C-term.

Table 1
GO functional classification of 358 genes selected from DNA microarray analysis.

Functional category (GO: molecular function)		Number of matched genes			
		Selected genes (358 genes)	%	RefSeq (24 210 cDNAs)	%
Binding	Nucleic acid binding	25	(14.4)	706	(8.9)
	Nucleotide binding	20	(11.5)	1047	(13.2)
	Ion binding	12	(6.9)	245	(3.1)
	Protein binding	9	(5.2)	450	(5.7)
	Other bindings	3	(1.7)	304	(3.8)
Catalytic activity	Transferase activity	19	(10.9)	808	(10.2)
	Hydrolase activity	15	(8.6)	902	(11.3)
	Ligase activity	9	(5.2)	107	(1.3)
	Other catalytic activity	11	(6.3)	653	(8.2)
Transcription regulator activity		18	(10.3)	364	(4.6)
Signal transducer activity		12	(6.9)	879	(11.0)
Transporter activity		11	(6.3)	997	(12.5)
Structural molecule activity		4	(2.3)	163	(2.0)
Enzyme regulator activity		3	(1.7)	142	(1.8)
Others		3	(1.7)	193	(2.4)
Total		174		7960	

Total refers to the number of cDNAs used for the classification of molecular function. We categorized each cDNA used for identifying the genomic regions of extracted 423 probes. Results obtained using the cDNAs transcribed from the same genomic region were merged. We categorized 24 210 human RefSeq and identified 20 072 protein-coding genes. Results obtained using the human RefSeq transcribed from the same genomic region were also merged. If an encoded protein was predicted to belong to two or more categories, it was counted every time.

We designed specific primers to detect protein-coding transcript, and succeeded in making specific primers for 30 genes (Supplementary Table 4). Using these specific primers, we then performed quantitative real-time PCR analysis to examine the expression profiles of the protein-coding transcripts produced by each one of the 30 genes (Supplementary Table 5). Results obtained for 4 genes in the "transcription regulator activity" category are described in Section 3.3 and results for the remaining genes are summarized in Supplementary Table 6(further detailed results are described in Supplementary analysis results). In the case of 12 genes, the expression profiles of the protein-coding transcripts produced by each one of them were different. The RA-induced change in the expression level of the soluble-form protein encoding transcript s9NRP2 (NM_201264.1) [18] was smaller than the change in the expression levels of the membrane-form protein encoding transcripts NRP2a (NM_003872.2) and NRP2b (NM_018534.3) (Supplementary Fig. 5A). In the case of erythrocyte membrane protein band 4.1 like 5 (EPB41L5), the expression of NM_020909.2 transcript, which encodes a protein known to bind to paxillin through its C-terminus [23], was up-regulated, whereas the expression level of BC032822.2 transcript, which encodes for a protein lacking the C-terminus paxillin binding site, did not change (Supplementary Fig. 5B). In the case of collagen triple helix repeat containing 1 (CTHRC1), the expression level of NM_138455.2 transcript, which encodes for a protein that is known to act as a coreceptor in the Wnt-PCP pathway [24], increased by 32-fold, whereas the expression level of FLJ57590 transcript, which encodes a protein lacking the C-terminus binding domain, increased by only fourfold after 7-day RA induction (Supplementary Fig. 5C). Interestingly, the expression level of NM_138455.2 decreased by twofold, whereas the expression level of FLJ57590 increased by 11-fold after 35-day of incubation with RA (Supplementary Fig. 6A). The transcripts produced by the nephronophthisis 1 (juvenile) gene, NM_000272.2 and NM_207181.1, showed similar expression patterns until day 7, and then the expression level of NM_207181.1 decreased, while the expression level of NM_000272.2 increased (Supplementary Fig. 6B). Notably, out of the 14 genes that showed similar expression patterns until 7-day, we found 7 genes in which the up or down-regulation ratios between the two transcripts produced from each gene were different at a later stage (Supplementary Table 6). Among these genes, there were two genes with unknown function, namely WD repeat domain 74 (Supplementary Fig. 6C) and DENN/MADD domain containing 5B (Supplementary Fig. 6D).

Next, we analyzed the relationship between the mRNA diversity and function of genes by utilizing the results of AS typing and GO classification (Supplementary Table 7). All genes in the "ligase activity" category produced multiple protein-coding transcripts. We predicted that multiple protein-coding transcripts having variable regions were required for neurodifferentiation. Ten genes in the "ion binding" category produced multiple protein-coding transcripts, 6 of which had alternative N-terminus, suggesting that these genes were controlled by transcription mechanism. The mRNA diversity of 18 genes in the "transcription regulator activity" category were analyzed in more detail (see below). We believe the mRNA diversity and gene function are correlated [25], and therefore, it is imperative to understand this relationship in order to identify the genes that are specifically involved in neurodifferentiation.

3.3. Analysis of expression patterns of selected transcription factors

We analyzed the mRNA diversity of 18 transcription factor genes which were classified under "transcription regulator activity" by GO analysis (Table 1, Supplementary Fig. 3), as transcription factors regulate various genes, and transcription factor genes producing multiple transcripts are known to work mutually as regulating factors in transcription [26]. Out of the 18 transcription factor genes examined, we found that 12 genes - namely ets variant 1 (ETV1), ets variant 4 (ETV4), ets variant gene 5 (ETV5), HNF1 homeobox B (HNF1B), homeobox A2 (HOXA2), homeobox A3 (HOXA3), paired box 6 (PAX6), paternally expressed 3 (PEG3), POU class 5 homeobox 1 (POU5F1), RARB, regulatory factor X, 2 (influences HLA class II expression) (RFX2) and zinc finger protein 483 (ZNF483) – produced multiple protein-coding transcripts by AS (Supplementary Fig. 7). Using primers designed against the specific regions of protein-coding transcripts (Supplementary Table 4), we performed quantitative real-time PCR analysis to examine the expression profiles of the protein-coding transcripts produced by each one of the 12 transcription factor genes (Supplementary Table 8). The results are summarized in Table 2, and results obtained on 6 of these genes are discussed below (further detailed results are described in Supplementary analysis results).

Table 2
Quantitative evaluation of 12 transcription factors by real-time PCR.

AS pattern	Gene symbol	cDNA name	Change in expression level compared to the 0-day sample			Expression pattern
			1-day	2 day	7 day	
Alt. N-terminus	ETV1	NM_004956.3	±	↓	±	Differ
		FLJ50494	±	±	↓	
	ETV4	NM_001986.1	±	↓	↓	Same
		BC007242.1	±	↓	↓	
	ETV5	NM_004454.1	↓	↓	↓	Differ
		FLJ56169	±	↓	↓	
	HOXA3	NM_030661.3	↑↑	↑↑↑	↑↑↑	Differ
		NM_153632.1	±	↑	↑	
	PAX6	NM_000280.2	↑	↑	↑↑	Same
		FLJ36930	↑	↑	↑↑	
	POU5F1	NM_002701.3	±	±	↓	Differ
		NM_203289.3	±	↓	↓	
	RARB	NM_000965.2	↑	↑↑	↑↑	Differ
		FLJ56241	↑	↑	↑	
Alt. C-terminus	HNF1B	NM_000458.1	↑	↑↑	↑↑	Same
		NM_006481.1	↑	↑↑	↑↑	
	HOXA2	NM_006735.3	↑↑↑	↑↑↑	↑↑↑	Differ
		FLJ39423	↑↑	↑↑	↑↑↑	
	PEG3	NM_006210.1	±	±	↑	Differ
		NM_015363.3	±	±	±	
	ZNF483	NM_133464.1	±	↓	↓	Differ
		FLJ35492	±	±	↓	
Alt. cassette-exon	RFX2	NM_000635.2	±	↑	↑	Same
		FLJ53376	±	↑	↑	

Out of the 18 transcription factors, 6 transcription factors – namely GATA6 (NM_005257.3), GBX2 (NM_001485.2), HESX1 (NM_003865.1), HOXB2 (NM_002145.2), HOXB4 (NM_024015.3) and HOXC4 (NM_014620.2) – produced only single protein-coding transcripts. They are not shown in this table. Symbols used to indicate differences of mean averages of log 2 ratios between the control (0-day) and experimental samples (1-day, 2-day and 7-day): "±", >1/2 but <twofold; "↓", >1/100 but ⩽1/2; "↑", ⩾twofold but <100-fold; "↑↑", ⩾100-fold but <1000-fold; and "↑↑↑", ⩾1000-fold.

In the case of PEG3 (Fig. 2A), HOXA2 (Fig. 2B) genes, we found that the regulation of mRNA diversity was controlled by the expressions level. Transcript NM_006210.1 encodes for PEG3 and transcript NM_015363.3 encodes for ZIM2 originated from one genome locus [27]. The expression of NM_006210.1 (PEG3) was preferentially up-regulated than the expression of NM_015363.3 (ZIM2) (Fig. 2A), suggesting that NM_006210.1 (PEG3) might be more necessary for neuronal differentiation. Since the tumor suppressor activity of PEG3 was shown to be due to the loss of cell proliferation of cancer cells [28], its expression level is expected to increase during the RA-induced neuronal differentiation of NT2 cells. Thus, our results suggest that protein-coding transcripts produced by a gene is regulated by AS. Likewise, the up-regulation ratios of the two transcripts, HOXA2 (NM_006735.3) and FLJ39423, were significantly different (Fig. 2B). FLJ39423, which was identified as the splicing variant of HOXA2 resulting from Alt. C-term utilization, was predicted to lack the homeobox domain (Supplementary Fig. 7I). These results suggest preferential need for the transcript containing functional domain during neuronal differentiation. We also obtained similar results in the case of RARB gene (Supplementary Table 8).

As shown in Fig. 2C, RA-induced increase in the expression levels of two transcripts, NM_000458.1 and NM_006481.1, produced from the HNF1B gene by utilizing Alt. C-terms were very similar. Similar results were observed with PAX6 and RFX2 genes (Supplementary Table 8), suggesting that both transcripts form each one of these genes might be needed for the neuronal differentiation, and they all were probably regulated by the same transcriptional regulatory mechanism.

From the expression profiles of two transcripts of the POU5F1 gene (Fig. 2D), resulting from the Alt. N-term utilization – NM_002701.3 known as OCT4A and NM_203289.3 known as OCT4B [29], we also found that the regulation of mRNA diversity was controlled both by the expressions level and the transcription time. The expression level of NM_203289.3 (OCT4B) decreased more rapidly than that of NM_002701.3 (OCT4A). We suggest that the decrease in the expression level of OCT4B was more important for the early stage of neuronal differentiation because its down-regulation was induced first. Similar results were observed with the ZNF483 gene (Supplementary Table 8). Since the unique regions of NM_002701.3 (OCT4A) and NM_203289.3 (OCT4B) chosen for the expression analysis were located only in their N-terminus, we suggest that the expression levels of both transcripts were regulated by different transcription mechanisms during neurodifferentiation.

Finally, we report on two other genes, HOXA3 and ETV5, of unknown functions. As shown in Fig. 2E, the up-regulation ratios and the time required to initiate changes in the expression level were different for two transcripts produced from the HOXA3 gene, NM_030661.3 and NM_153632.1, which differed in their N-terminal regions. We speculated that NM_030661.3 played a more important role than NM_153632.1 in neuronal differentiation. We next examined expression of ETV5 (NM_004454.1) and FLJ56169, the latter of which was previously identified as the splicing variant of the former [11], and they differ in the N-terminus as a result of first exon variation. The expression level of NM_004454.1 decreased rapidly than the expression level of FLJ56169 (Fig. 2F). Interestingly, the expression level of NM_004454.1 increased back to the control (0-day) level in the 35-day sample (Supplementary Table 8) whereas the expression level of FLJ56169 increased back to the control level after 14-day and then increased by fourfold after 35-day incubation with RA. Based on these results, we suggest that whereas a decrease in the expression level of NM_004454.1 was needed at an early stage of neuronal differentiation, an increase in the expression level of FLJ56169 was needed at a later stage. In the case of ETV1 gene (Supplementary Table 8), we obtained similar result as ETV5 gene. The transcripts produced by the ETV4 gene showed similar expression patterns as ETV5 gene until day 7, and then the expressions of the ETV4 transcripts were up-regulated differently than those of

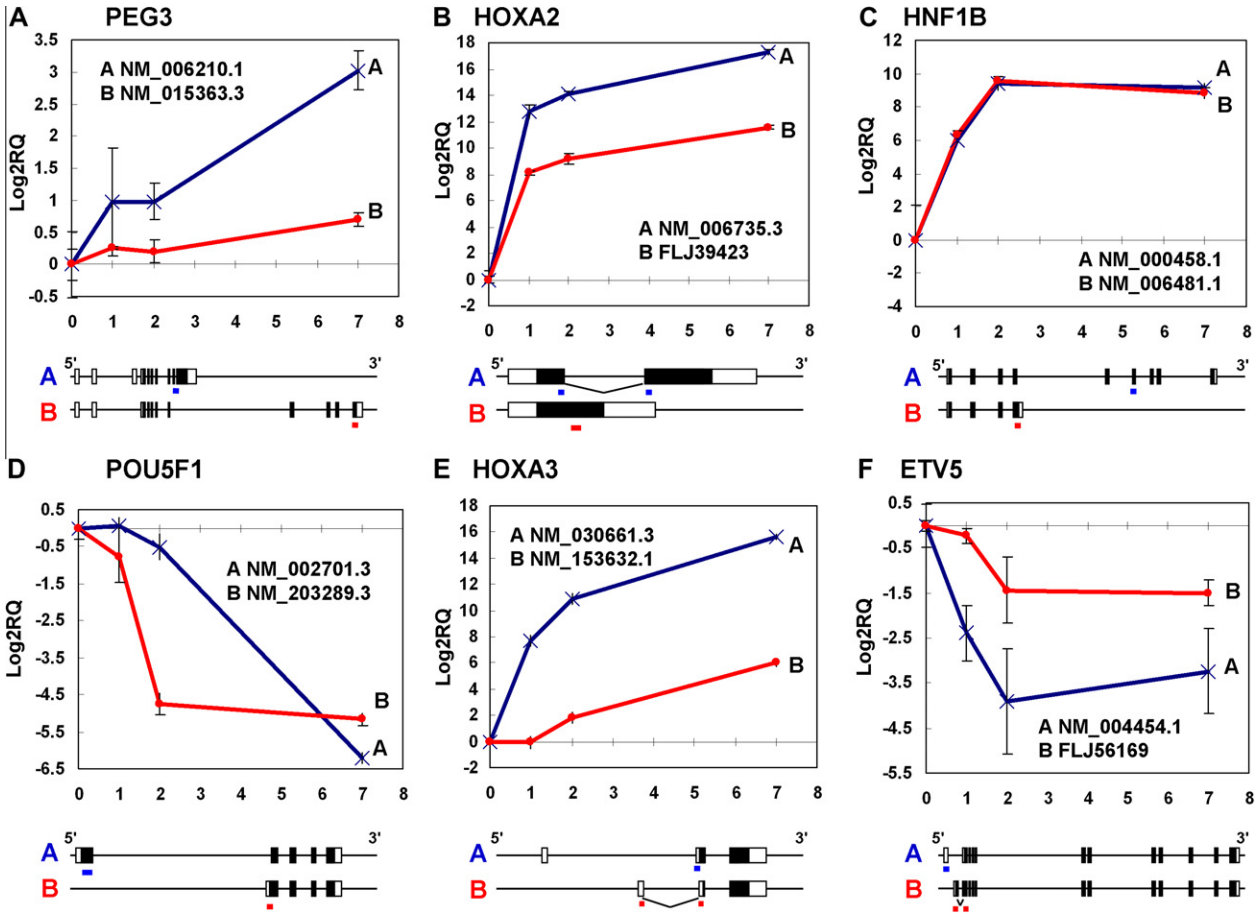

Fig. 2. Quantitative evaluation of selected 6 transcription factors by real-time PCR. Expression levels of two transcripts produced by AS from each gene were analyzed by real-time PCR, and were represented in log 2 base. The data were normalized with respect to that of the human GAPDH. Name of the genes: (A) PEG3, (B) HOXA2, (C) HNF1B, (D) POU5F1, (E) HOXA3, (F) ETV5. Schematic views of selected genes: boxes, exons; gray lines, introns; black boxes, coding regions; white boxes, untranslated regions; red or navy bars, amplified regions by real-time PCR.

the ETV5 transcripts (Supplementary Table 8). Thus, we found that the transcripts produced by the members of the ETS domain family, the ETV5, ETV4 and ETV1 genes, showed similar expression patterns.

4. Discussion

To know the underlying causes of diseases related to neuronal differentiation, we analyzed the expression patterns of genes in NT2 cells that were induced by RA to differentiate into neurons [12]. Consequently, we identified 358 RA-responsive genes following 7-day induction with RA. GO classification of selected genes revealed three major functional groups (Table 1). Eighteen genes belonging to the "transcription regulator activity" category were identified as transcription factors. Most genes under the categories, "ligase activity" and "ion binding", were related to cell properties (Supplementary Table 3).

AS of a gene results in multiple protein-coding transcripts [1–3]. AS particularly occurs and plays an important role in imparting biological and functional diversity associated with neuronal differentiation [4,5]. To understand functions of the RA-responsive genes, it was necessary to analyze their mRNA diversity. Our analysis using the FLJ Human cDNA Database ver. 3.0 [11] revealed that about 77% of our selected genes produced multiple protein-coding transcripts by AS (Supplementary Fig 4 and Supplementary Table 2). We have previously used the FLJ Human cDNA Database to reliably evaluate mRNA diversity of genes and correlated mRNA diversity with gene function.

We examined what effect neuronal differentiation has on the expression of alternatively spliced transcripts of the 18 RA-responsive transcription factor genes (Table 2). These transcription factor genes are expected to play key roles in neuronal differentiation. In PEG3 (Fig. 2A), HOXA2 (Fig. 2B) and RARB (Supplementary Table 8) genes, expression of one of the two transcripts was preferentially up-regulated than the other one, suggesting that they might be required for neurodifferentiation. On the other hand, RA similarly induced the expression of two transcripts from each one of the HNF1B (Fig. 2C), PAX6, ETV4 and RFX2 (Supplementary Table 8) genes, this observation predicts that both transcripts were needed in neurodifferentiation. Further studies are needed to confirm whether these two transcripts are indeed play any role in neurodifferentiation. In the case of POU5F1 (Fig. 2D), HOXA3 (Fig. 2E), ETV5 (Fig. 2F), ETV1 and ZNF483 (Supplementary Table 8) genes, not only the up-regulation ratios of two transcripts produced from each gene, but also the time required to initiate changes in their expression level were different. We speculate that the mechanism by which these genes could selectively produce a given transcript might be related to neurodifferentiation. Especially in HOXA3 (Fig. 2E) and ETV5 (Fig. 2F) genes, because the two transcripts produced by each one of these two genes differed only in their N-terminus, it strongly suggested that their N-terminus regions have some yet to be determined key function in neuronal differentiation. Since these N-terminus ends also included TSS, we suggest that their specific expression were controlled by transcriptional mechanism. Thus, by comparing expression profiles of different transcripts, we obtained new functional information on genes.

We also analyzed the expression profiles of genes that produced multiple protein-coding transcripts by using alternate C-terminus. In some cases, we found that the RA-induced changes in the expression of transcripts produced by the gene were different (Supplementary Table 6). Genes included in this category were some of the disease-related genes – for example, NRP2 gene thought to be related to tumor progression [18,30], and EPB41L5 gene is thought to be related to cancer cell invasion [23]. These results suggested that any approach that includes multiple protein-coding transcripts in the analysis process would be important for furthering functional analysis of genes. We also found that in the case of the CTHRC1 gene, the expression level of the transcript, which codes for a factor in the Wnt-PCP pathway [24], was induced at the early stage, whereas the expression level of FLJ57590 transcript that lacks the C-terminus binding site for the Wnt-PCP pathway factor was induced at a later stage following RA induction (Supplementary Fig. 6A). We think that the protein encoded by the FLJ57590 transcript is most likely not related to the Wnt-PCP pathway, but might be related to the canonical Wnt-signaling pathway.

Not only genetic backgrounds but environmental conditions might contribute to the mRNA diversity. Identification of the risk factor responsible for the mRNA diversity would, therefore, be an important step for discovering new drugs. Results described in this study tend to suggest that the mRNA diversity of genes might play important roles in neuronal differentiation induced by the environmental conditions, such as RA. Our study also suggest that by investigating how genes are alternatively spliced during neuronal differentiation might contribute significantly for understanding of the mechanism of neuronal differentiation. Accumulating more data on the mRNA diversity of genes using approaches similar in this study will not only reveal the underlying mechanism by which genes produce specific protein-coding transcripts in response to various factors but will also contribute to the uncovering of new functions of genes. Additionally, more and better we understand AS and its contribution to diseases, we might be able to identify candidate genes as novel targets for new drugs.

Acknowledgements

This work was partly supported by a Grant from the New Energy and Industrial Technology Developmental Organization (NEDO) project of the Ministry of Economy, Trade and Industry of Japan. We thank the members of the NEDO functional analysis of protein and research application project for generating the basal data and creating the database, and we specially thank K. Kimura for the genome mapping.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2010.08.024.

References

[1] Lander, E.S. et al. (2001) Initial sequencing and analysis of the human genome. Nature 409, 860–921.
[2] Wang, E.T. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. Nature 456, 470–476.
[3] Stamm, S. (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. Hum. Mol. Genet. 11, 2409–2416.
[4] Li, Q. et al. (2007) Neuronal regulation of alternative pre-mRNA splicing. Nat. Rev. Neurosci. 8, 819–831.
[5] Boutz, P.L. et al. (2007) A post-transcriptional regulatory switch in polypyrimidine tract-binding proteins reprograms alternative splicing in developing neurons. Genes Dev. 21, 1636–1652.
[6] Ule, J. et al. (2005) Nova regulates brain-specific splicing to shape the synapse. Nat. Genet. 37, 844–852.
[7] Licatalosi, D.D. et al. (2008) HITS-CLIP yields genome-wide insights into brain alternative RNA processing. Nature 456, 464–469.
[8] Ota, T. et al. (2004) Complete sequencing and characterization of 21, 243 full-length human cDNAs. Nat. Genet. 36, 40–45.
[9] Otsuki, T. et al. (2005) Signal sequence and keyword trap in silico for selection of full-length human cDNAs encoding secretion or membrane proteins from oligo-capped cDNA libraries. DNA Res. 12, 117–126.
[10] Kimura, K. et al. (2006) Diversification of transcriptional modulation: large-scale identification and characterization of putative alternative promoters of human genes. Genome Res. 16, 55–65.
[11] Wakamatsu, A. et al. (2009) Identification and functional analyses of 11, 769 full-length human cDNAs focused on alternative splicing. DNA Res. 16, 371–383.
[12] Andrews, P.W. (1984) Retinoic acid induces neuronal differentiation of a cloned human embryonal carcinoma cell line in vitro. Dev. Biol. 103, 285–293.
[13] Miura, A. et al. (2006) Differential responses of normal human coronary artery endothelial cells against multiple cytokines comparatively assessed by gene expression profiles. FEBS Lett. 580, 6871–6879.
[14] Tollervey, D. (2006) Molecular biology: RNA lost in translation. Nature 440, 425–426.
[15] Houldsworth, J. et al. (2002) Expression profiling of lineage differentiation in pluripotential human embryonal carcinoma cells. Cell Growth Differ. 13, 257–264.
[16] Balmer, J.E. and Blomhoff, R. (2002) Gene expression regulation by retinoic acid. J. Lipid Res. 43, 1773–1808.
[17] Levy, E. et al. (1987) Structure and evolutionary origin of the gene encoding mouse NF-M, the middle-molecular-mass neurofilament protein. Eur. J. Biochem. 166, 71–77.
[18] Rossignol, M. et al. (2000) Genomic organization of human neuropilin-1 and neuropilin-2 genes: identification and distribution of splice variants and soluble isoforms. Genomics 70, 211–222.
[19] Dehmelt, L. and Halpain, S. (2005) The MAP2/Tau family of microtubule-associated proteins. Genome Biol. 6, 204.
[20] Pavel, M.E. et al. (2006) Adrenomedullin is a novel marker of tumor progression in neuroendocrine carcinomas. Horm. Metab. Res. 38, 112–118.
[21] Landry, J.R. et al. (2003) Complex controls: the role of alternative promoters in mammalian genomes. Trends Genet. 19, 640–648.
[22] Zhang, T. et al. (2004) Multiple variable first exons: a mechanism for cell- and tissue-specific gene regulation. Genome Res. 14, 79–89.
[23] Hirano, M. et al. (2008) EPB41L5 functions to post-transcriptionally regulate cadherin and integrin during epithelial-mesenchymal transition. J. Cell Biol. 182, 1217–1230.
[24] Yamamoto, S. et al. (2008) Cthrc1 selectively activates the planar cell polarity pathway of Wnt signaling by stabilizing the Wnt-receptor complex. Dev. Cell. 15, 23–36.
[25] Sato, N. et al. (2001) Increased production of beta-amyloid and vulnerability to endoplasmic reticulum stress by an aberrant spliced form of presenilin 2. J. Biol. Chem. 276, 2108–2114.
[26] Laitem, C. et al. (2009) Ets-1 p27: a novel Ets-1 isoform with dominant-negative effects on the transcriptional properties and the subcellular localization of Ets-1 p51. Oncogene 28, 2087–2099.
[27] Kim, J. et al. (2000) Exon sharing of a novel human zinc-finger gene, ZIM2, and paternally expressed gene 3 (PEG3). Genomics 64, 114–118.
[28] Kohda, T. et al. (2001) Tumour suppressor activity of human imprinted gene PEG3 in a glioma cell line. Genes Cells 6, 237–247.
[29] Atlasi, Y. et al. (2008) OCT4 spliced variants are differentially expressed in human pluripotent and nonpluripotent cells. Stem Cells 26, 3068–3074.
[30] Geretti, E. and Klagsbrun, M. (2007) Neuropilins: novel targets for anti-angiogenesis therapies. Cell. Adh Migr. 1, 56–61.